

R Notebook

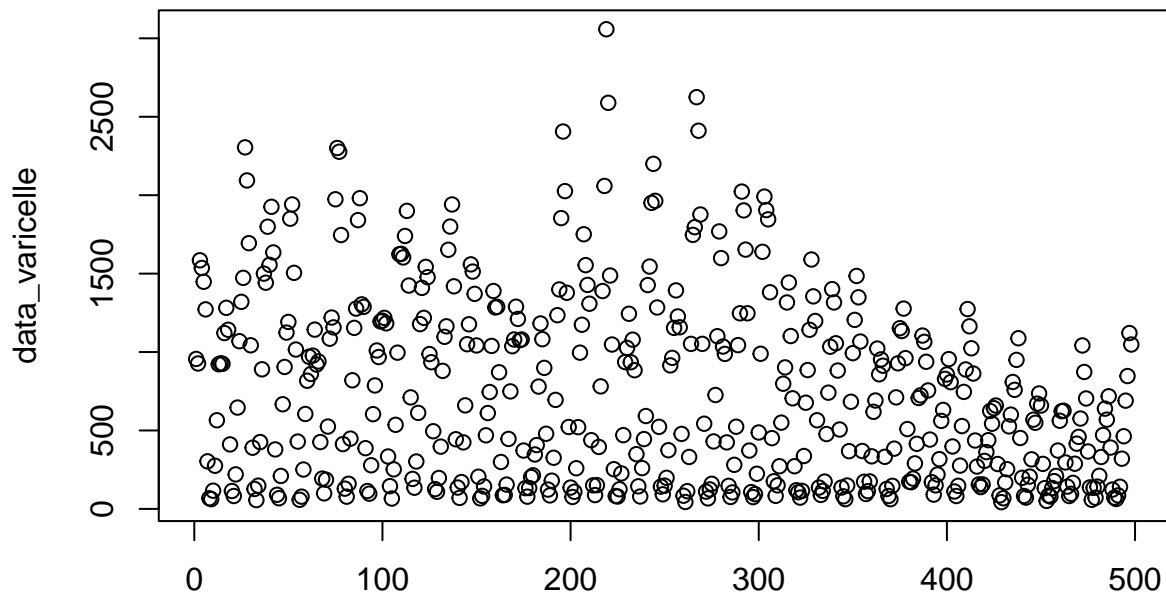
Mlamali SAID SALIMO

19 January, 2022, 16:01

```
data_varicelle = scan(file="data/varicelle.dat",skip=1)
head(data_varicelle)
```

```
## [1] 956 927 1585 1536 1448 1272
```

```
plot(data_varicelle)
```



Index

On

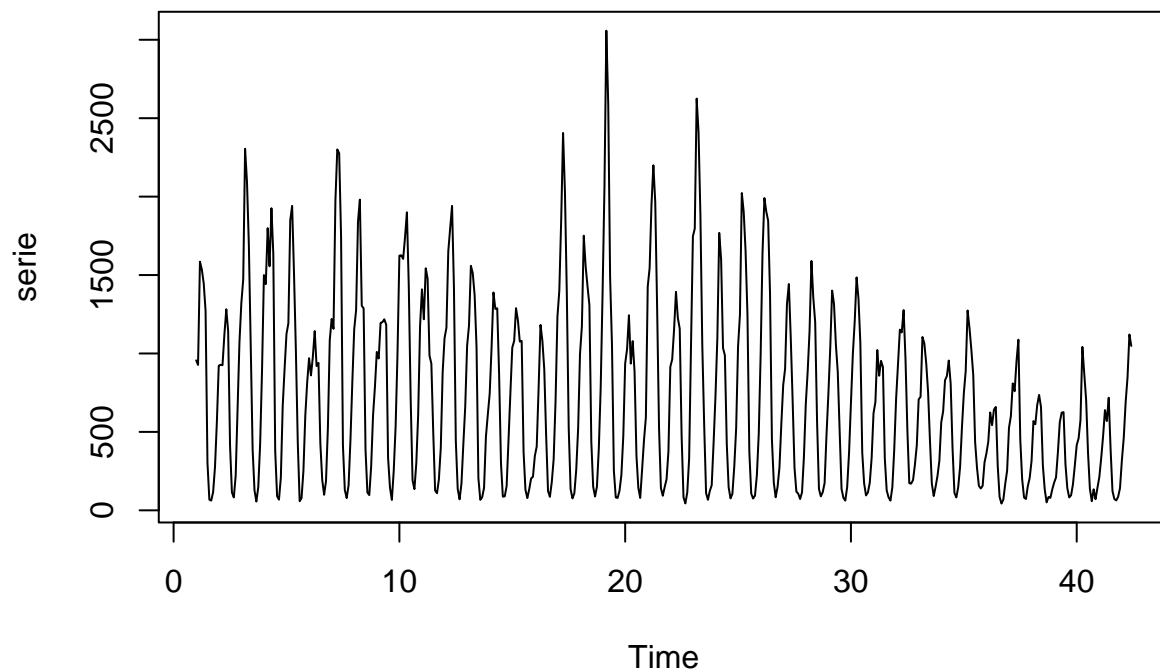
transforme les données en serie temporelles.

```
serie = ts(data_varicelle,frequency = 12) #frequence pour avoir les années en absciesse (les données me  
serie
```

##	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
## 1	956	927	1585	1536	1448	1272	303	68	62	116	275	565
## 2	922	928	925	1121	1282	1142	411	114	82	220	646	1069
## 3	1320	1473	2305	2094	1694	1043	390	127	56	148	426	890
## 4	1500	1442	1799	1556	1926	1635	379	90	68	210	667	905
## 5	1124	1192	1850	1941	1505	1016	429	58	78	251	605	817
## 6	970	860	977	1143	920	940	426	193	99	186	525	1085
## 7	1220	1157	1974	2301	2277	1746	413	129	78	160	448	820
## 8	1154	1277	1841	1981	1304	1288	387	114	97	278	604	787
## 9	1010	968	1195	1200	1218	1183	334	145	66	252	536	996
## 10	1624	1626	1603	1740	1900	1424	711	191	135	302	612	1178

```
## 11 1409 1218 1543 1477 987 935 495 126 109 197 397 880
## 12 1097 1164 1652 1800 1941 1419 444 136 70 171 424 660
## 13 1050 1177 1559 1513 1371 1042 205 67 83 143 469 611
## 14 745 1039 1389 1284 1288 871 299 87 89 155 446 749
## 15 1037 1080 1289 1211 1076 1080 372 132 78 133 203 214
## 16 347 407 780 1182 1082 899 479 123 86 180 326 695
## 17 1235 1399 1854 2406 2026 1378 522 136 76 109 259 521
## 18 996 1174 1751 1554 1428 1308 438 150 88 151 395 781
## 19 1389 2059 3058 2589 1488 1048 253 82 79 125 226 470
## 20 936 1026 1244 935 1079 884 349 144 79 260 445 592
## 21 1427 1545 1951 2200 1964 1284 523 142 93 148 198 374
## 22 915 963 1154 1393 1227 1158 478 84 44 113 331 1052
## 23 1747 1796 2625 2411 1877 1052 543 110 67 124 160 430
## 24 726 1101 1769 1599 1035 988 424 147 76 105 281 524
## 25 1044 1247 2023 1903 1653 1247 372 107 75 94 224 487
## 26 989 1639 1991 1905 1846 1381 451 176 83 150 272 550
## 27 798 902 1316 1443 1102 705 272 119 106 72 115 337
## 28 677 885 1142 1590 1355 1198 565 136 89 115 174 477
## 29 741 1034 1401 1316 1056 882 506 136 80 62 149 368
## 30 683 993 1205 1485 1349 1067 369 173 95 113 175 335
## 31 619 691 1022 858 953 913 332 127 82 62 147 384
## 32 711 928 1152 1134 1277 961 509 173 170 193 290 415
## 33 707 724 1105 1065 938 755 442 170 91 150 219 317
## 34 561 631 829 857 955 808 398 111 82 147 276 528
## 35 746 889 1274 1164 1024 863 436 270 156 139 156 306
## 36 362 438 624 543 642 659 286 86 43 68 168 253
## 37 526 601 809 759 950 1088 452 198 82 72 154 206
## 38 316 569 549 671 736 659 287 132 51 85 79 133
## 39 177 210 372 562 623 626 296 142 82 96 166 288
## 40 416 459 576 1042 873 704 366 137 58 134 71 142
## 41 211 331 471 639 569 718 391 123 72 63 86 141
## 42 320 463 690 847 1121 1048
```

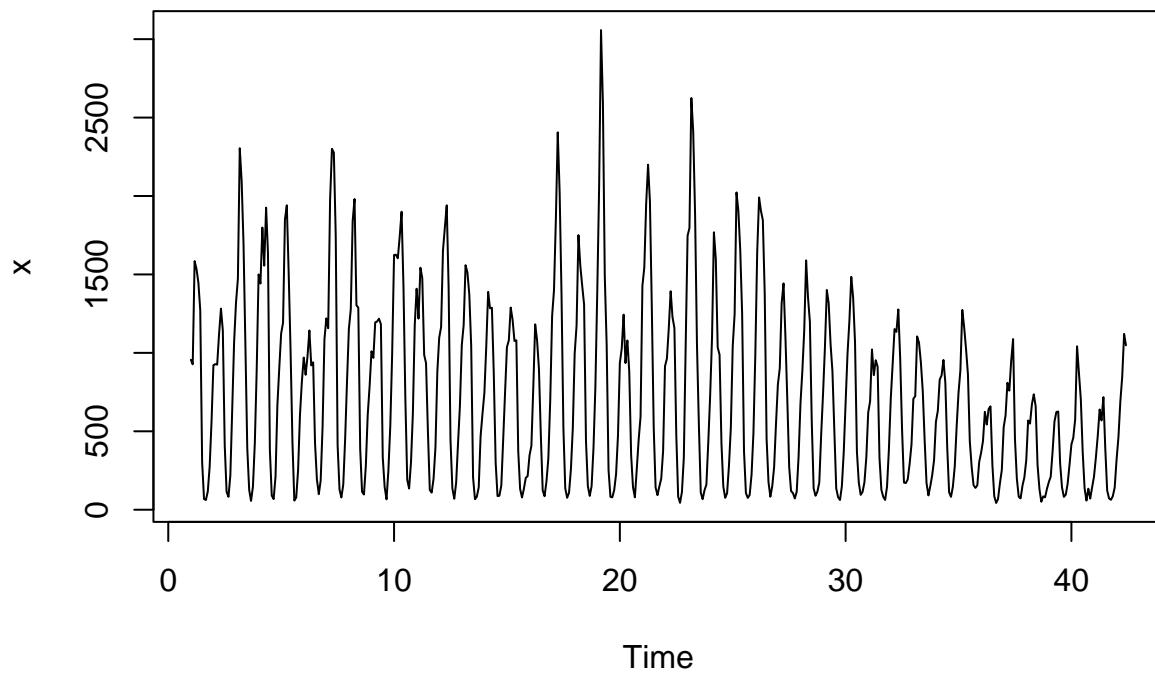
```
plot(serie) #dessiner le graphique
```



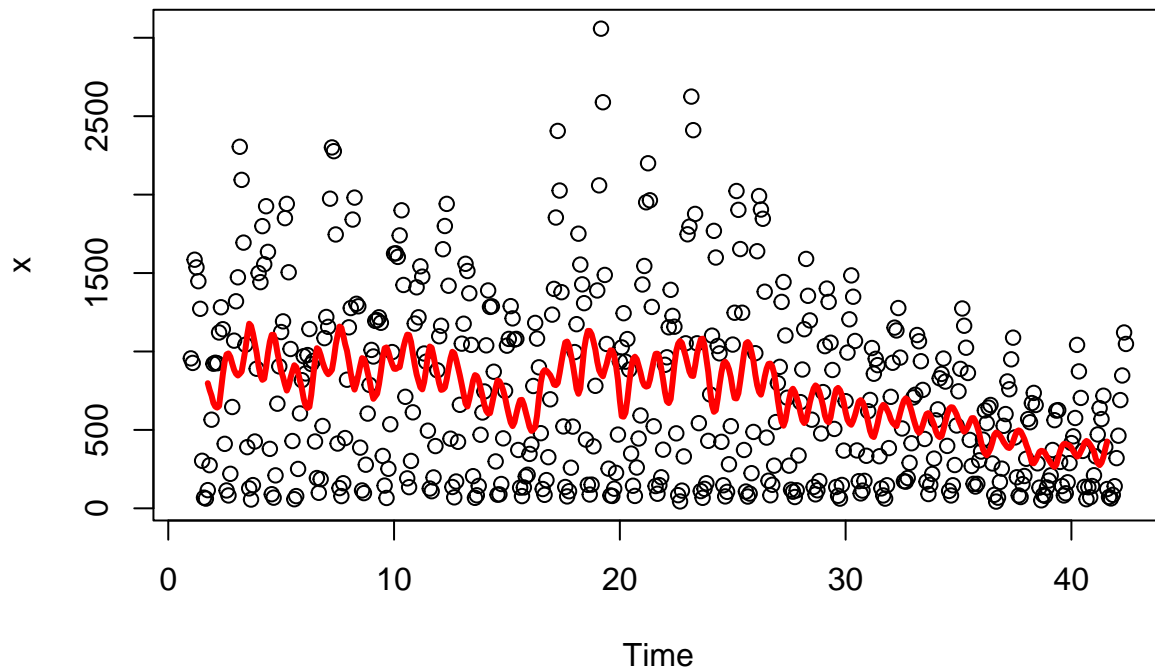
ne voit pas grand chose on peut lisser la courbe

si on

```
x = serie
plot(x)
```



```
plot(x,type='p')
k = 20
lines(filter(x,rep(1/k,k)), col='red', lwd=3)
```



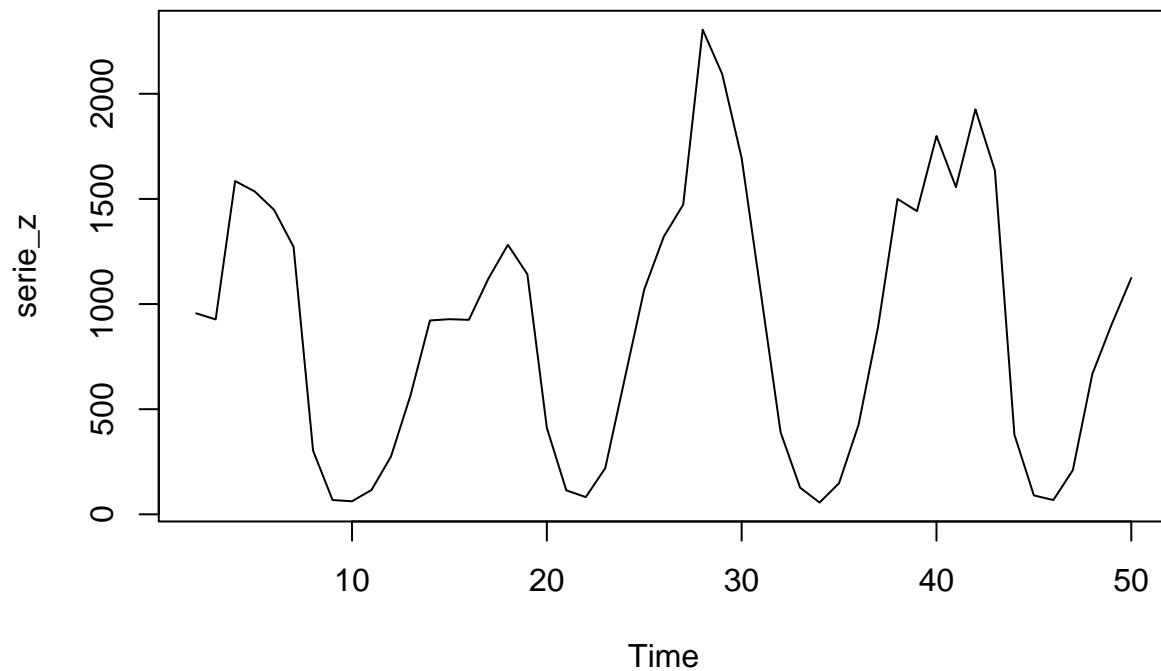
```
# prendre 20 valeur de x aléatoire
```

Question 2 - Analyser la série

On ne voit pas de tendance mais on voit un saisonnalité

Pour mieux voir la saisonnalité, on peut zoomer sur une durée plus petite. Les 50 premiers mois par exemple.

```
serie_z = ts(data_varicelle,2,50)
plot(serie_z)
```



Question 3 - calcul de la moyenne

```
mean(serie)
```

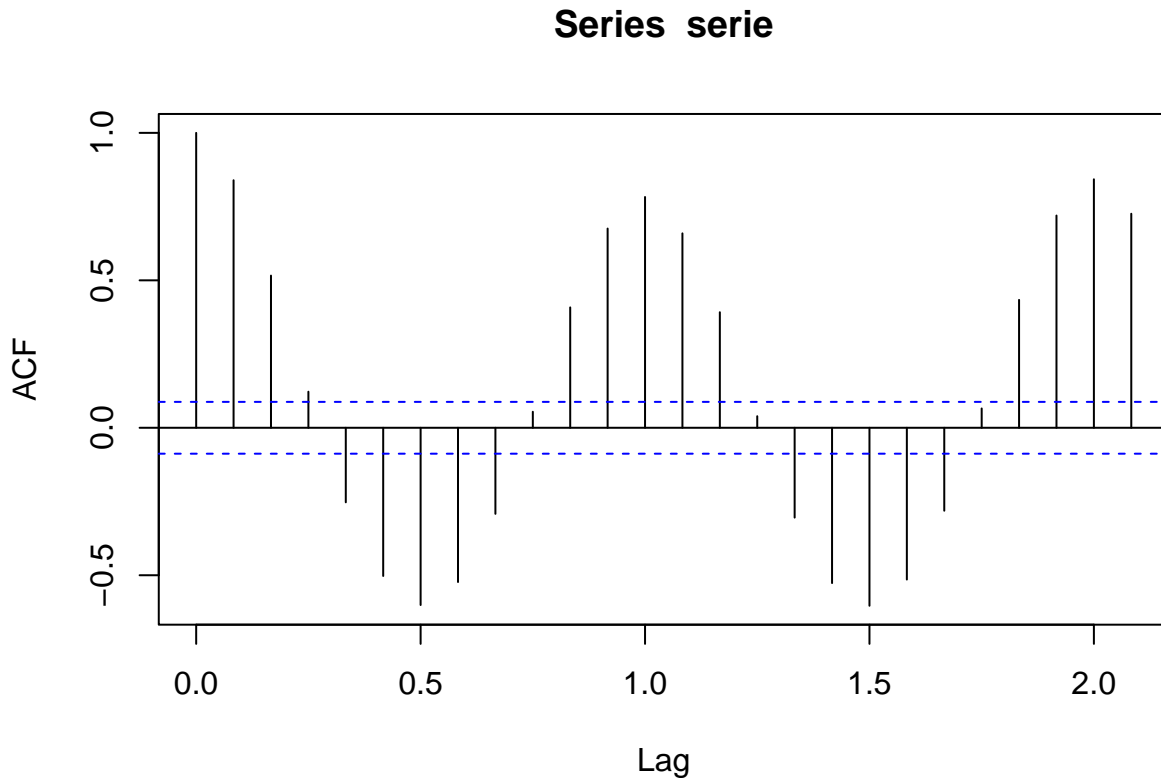
```
## [1] 732.4076
```

Question 4 - ACF des 25 premières auto-corrélations

Def et objectif Fct de auto-corrélation :

```
#On rentre l'instruction
```

```
acf(serie,lag.max=25,type=c("correlation"))
```



on voit que ça converge, mais lentement. Tout ce qui est à l'intérieur de l'invalle on le considère statistiquement nulle (non significatif). Elle converge tjr vers 0 chakal la juste lentement

- on observe une saisonnalité la saisonnalité observée graphiquement est confirmée par la fct de autocorrélation.

On a fait avec 25 valeur et sa suffit largement

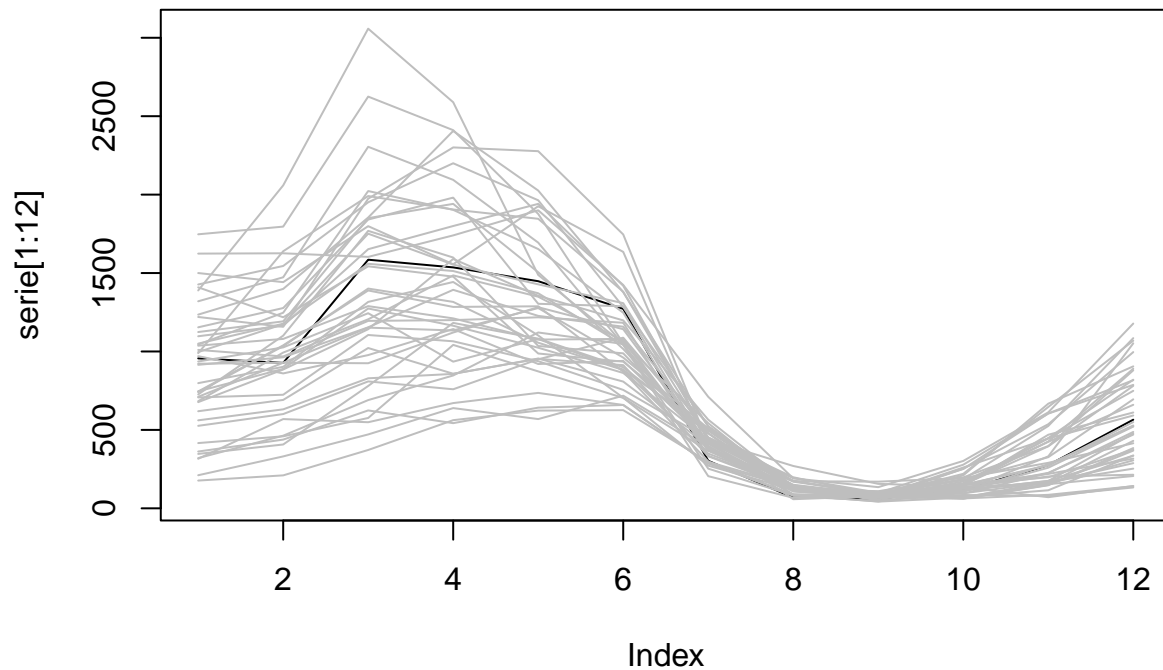
La fonction d'autocorrélation est périodique, ce qui indique une périodicité de la série temporelle.

La ligne pointillé bleu indique le niveau en dessous duquel la corrélation n'est plus statistiquement significative

Question 5 - Tracer sur un même graphique

Les évolutions mensuelles de cas varicelles pour chaque année

```
plot(serie[1:12],type="l",ylim=c(min(serie),max(serie)))  
for (i in 1 : 41) lines(serie[(1+12*i) : (12*(i+1))],col="grey")
```

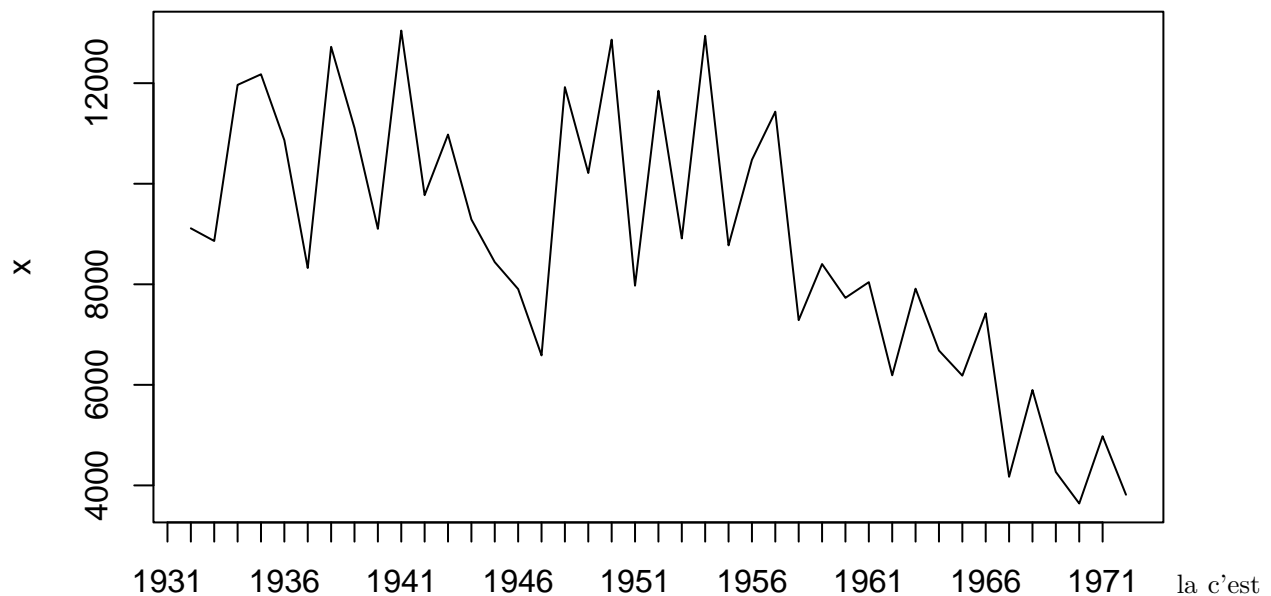


Question 6 : Tracer sur un graphique l..

```
x=rep(0,41)
for(i in 0 : 40) x[i+1] = sum(serie[(1+12*i) : (12*(i+1))])
x
```

```
## [1] 9113 8862 11966 12177 10866 8324 12723 11112 9103 13046 9773 10978
## [13] 9290 8441 7905 6586 11921 10214 12866 7973 11849 8912 12942 8775
## [25] 10476 11433 7287 8403 7731 8042 6190 7913 6683 6183 7423 4172
## [37] 5897 4267 3640 4978 3815
```

```
plot(x,type="l", xaxt='n', xlab='')
axis(1,at = 0 : 40,labels = 1931 : 1971)
```



évolution annuelles du nombre de cas de varicelle (tout à l'heure et l'évolution mensuelle pour chaque année)

Remarques : - a partir de 1955, la tendance de la diminution a été confirmée - ça c'est été maîtriser (la maladie) vers 1971

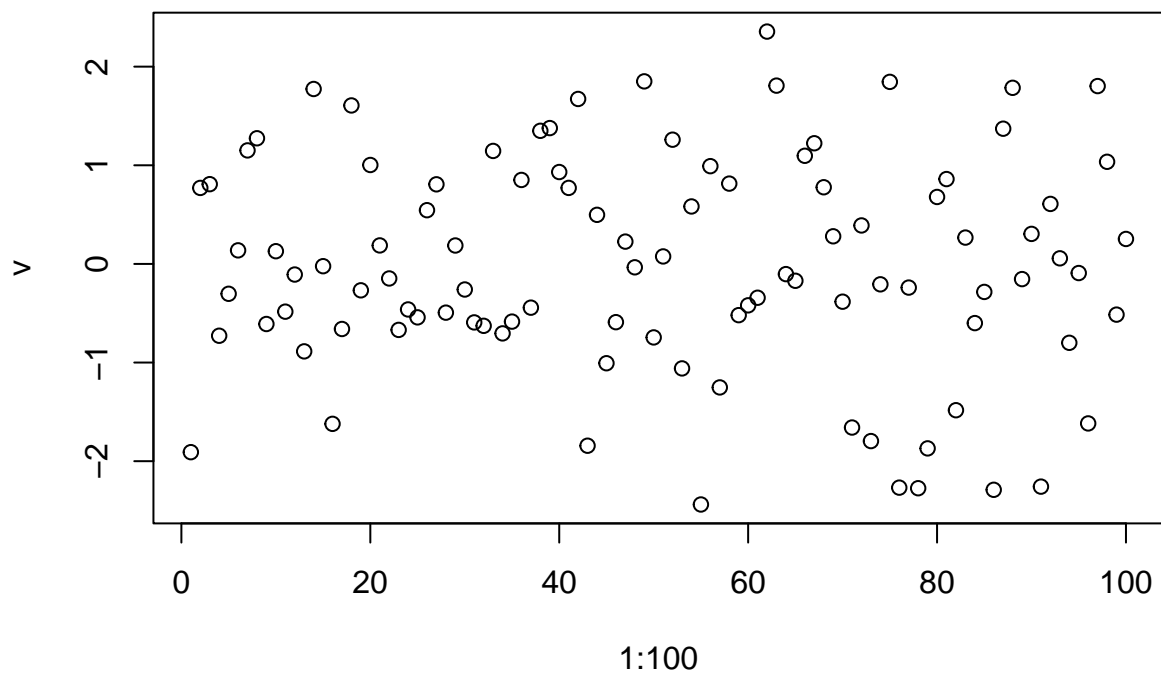
Exercice 2

1)

voir photo

2) Simuler un .

```
v = rnorm(100,0,1)
plot(1 : 100, v)
```

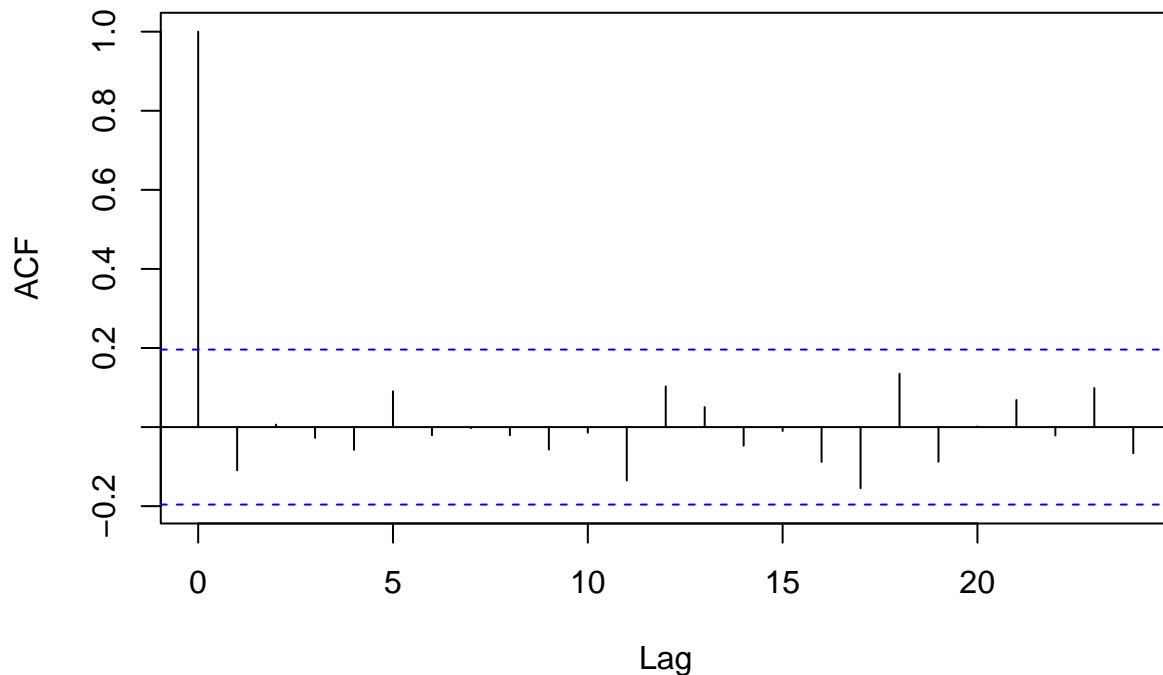


3)

traver la fct de auto

```
acf(v, lag.max = 24, type = c("correlation"))
```

Series v



- abs-
ciser : h, ordonnée $\rho(h)$ “normalement on devrai pas avoir plus de 5 valeur à l’exterieur de l’intervalle : psk
nanani 5% param rnorm” “les termes sont corrélarer sauf ρ de 0, donc le graphique reflète bien la pertie
théorique

y’a pas dévolution de la, sur le graphique tous les valeur de ρ de 1 sont tout statistiquement non signiicative
rappel : l’intervalle de confiance d’pende de n

4) Recommencer les deux question et oberser la variabilité

5) simuler la série temporelle

avant et un bruit blanc, mtn une serie avec tendance

```
v = rnorm(100,mean=0,sd=1)
t = 1:100

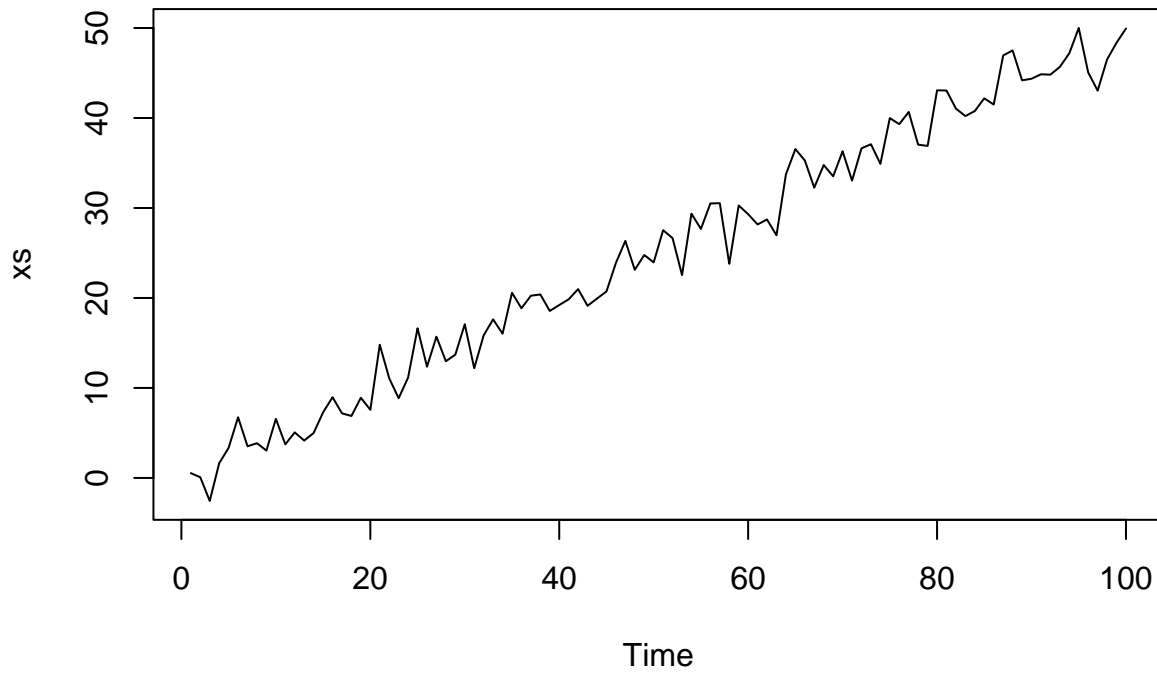
x = 0.5*t+2*v
xs = ts(x)
xs
```

```
## Time Series:
## Start = 1
## End = 100
## Frequency = 1
## [1] 0.53237954 0.07226997 -2.55188418 1.62136978 3.35528198 6.74142824
## [7] 3.51606406 3.86169051 3.04152666 6.56536995 3.73143975 5.06300402
## [13] 4.15341332 4.99425783 7.28855944 8.97143727 7.18183854 6.88287215
## [19] 8.91209973 7.57047983 14.79987639 11.07629851 8.85751393 11.15135053
## [25] 16.63112751 12.36528125 15.69452948 12.96580270 13.70530628 17.08194674
## [31] 12.19233159 15.84658819 17.62189062 16.03220546 20.57872571 18.85318557
## [37] 20.24725008 20.39330744 18.55643327 19.21423300 19.84845553 20.98791529
```

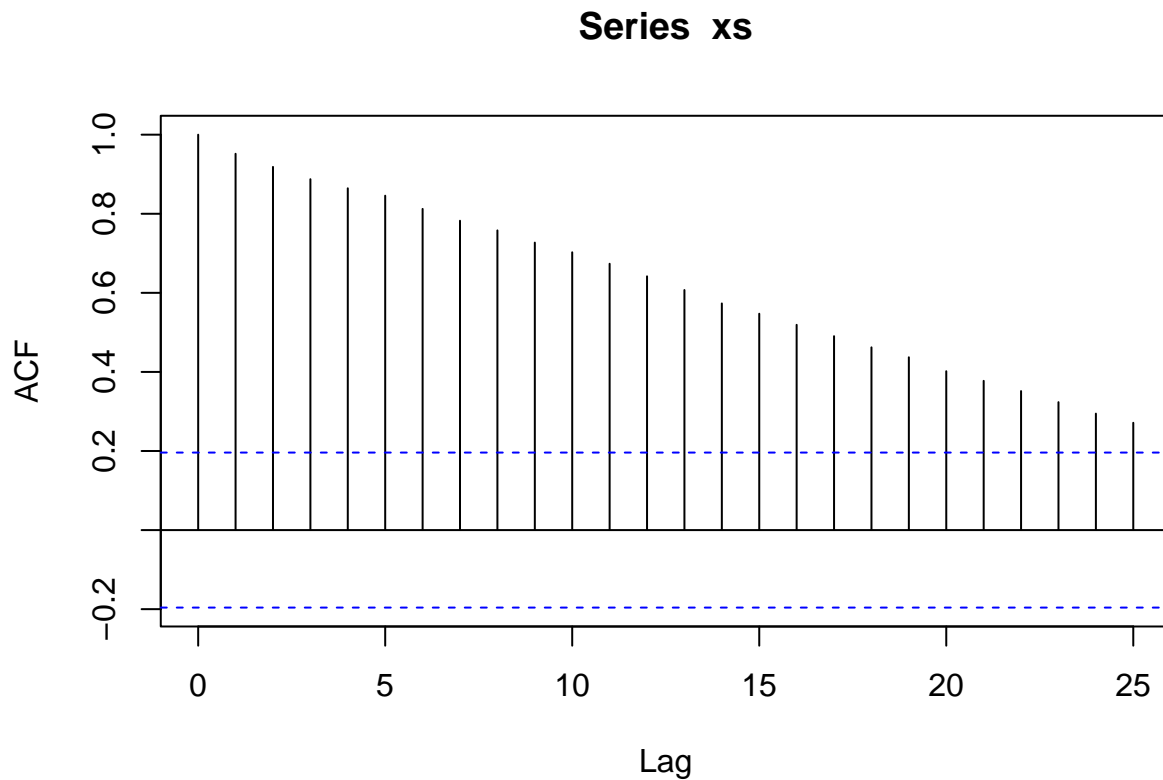


```
## [43] 19.12861526 19.93783514 20.72302983 23.89485798 26.34214624 23.12792573
## [49] 24.77987414 23.95367805 27.53282065 26.65092571 22.54598025 29.36669957
## [55] 27.66867354 30.50105438 30.53487738 23.79428971 30.28475175 29.30824170
## [61] 28.17437491 28.72192882 26.96802100 33.74320798 36.54654938 35.28019039
## [67] 32.26065173 34.77736345 33.51793950 36.30214706 33.04208476 36.62852194
## [73] 37.08568044 34.90518195 39.99084336 39.31744789 40.67135561 37.03680127
## [79] 36.89370249 43.07829119 43.05283196 41.03737294 40.21382539 40.77586673
## [85] 42.19145218 41.49010309 46.94920585 47.50686472 44.18638817 44.36306813
## [91] 44.85940396 44.81749933 45.68059129 47.19131273 50.00181696 45.06582644
## [97] 43.04453136 46.50746493 48.36440416 49.94257833
```

```
plot(xs)
```



```
acf(xs,lag.max = 25,type = c("correlation"))
```



Suite à tendance croissante en général

$$X_t = 5t + \text{Eps}_t$$

$$E(X_t) = 5t$$

la serie X_t n'est pas stationnaire, ya une tendance (la moyenne n'est pas constante). Cette tendance ça se voit graph de fct de autocorrélation : on observe une décroissance lente de quel converge vers 0 mais très lentement

Q7

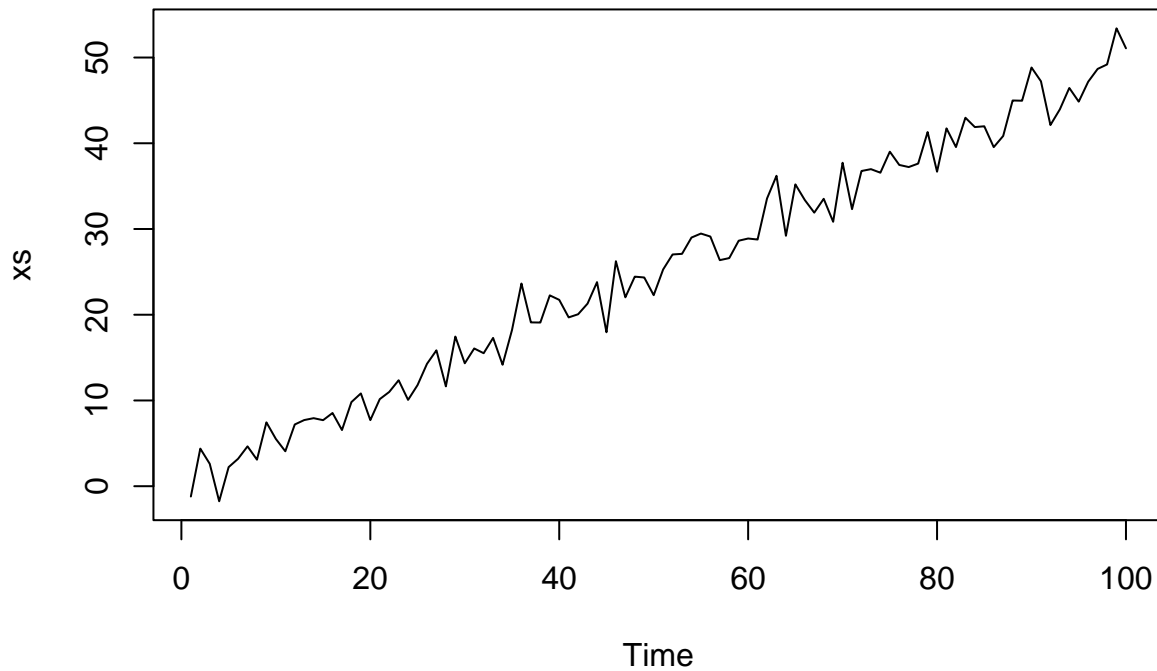
```
t = 1:100
v = rnorm(100,mean=0,sd=1)

x = 0.5*t+2*v
xs = ts(x)
xs
```

```
## Time Series:
## Start = 1
## End = 100
## Frequency = 1
## [1] -1.204183  4.385621  2.612780 -1.755320  2.226432  3.198629  4.646286
## [8]  3.095665  7.446487  5.513076  4.071751  7.197682  7.715825  7.934797
## [15]  7.702701  8.539142  6.550395  9.810910 10.822686  7.709293 10.155609
## [22] 10.991156 12.359128 10.064820 11.794602 14.273851 15.848281 11.645859
## [29] 17.452566 14.336354 16.066539 15.511782 17.298612 14.169478 18.219511
## [36] 23.637994 19.107970 19.088555 22.258339 21.734826 19.688454 20.056672
## [43] 21.296113 23.801713 17.957397 26.242857 22.031998 24.441594 24.344272
## [50] 22.281015 25.277493 27.028715 27.101314 29.001794 29.468987 29.120257
```

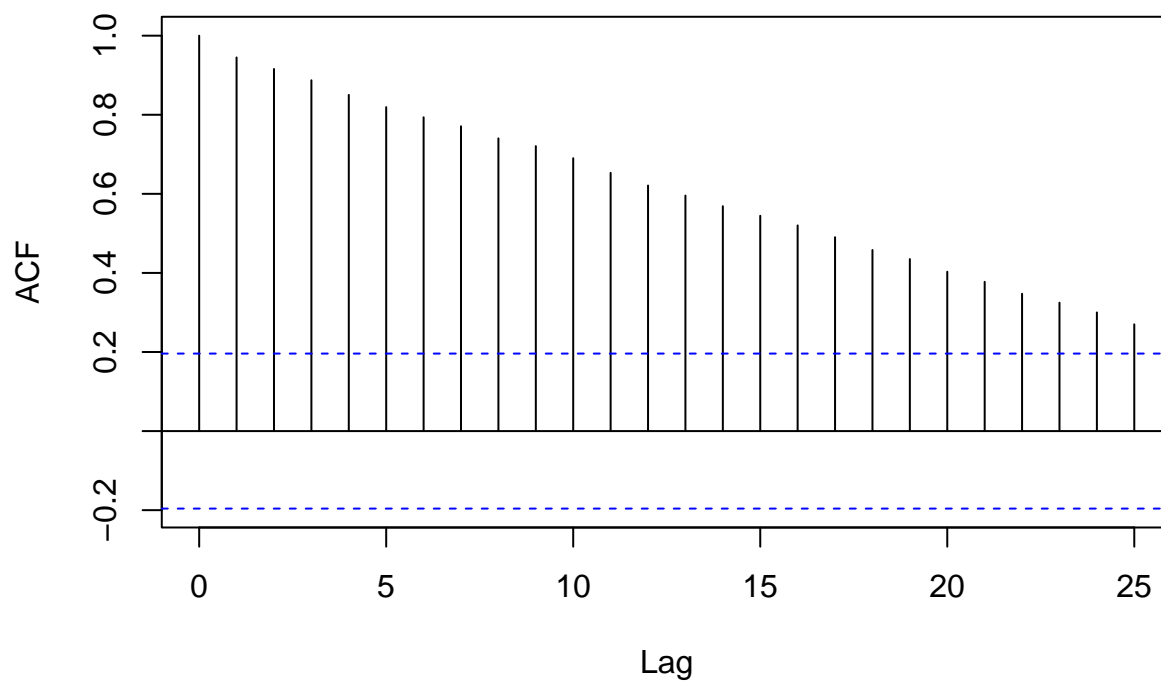
```
## [57] 26.367510 26.607048 28.634228 28.889281 28.771179 33.558832 36.205941
## [64] 29.208685 35.203976 33.376491 31.905907 33.522934 30.843207 37.726184
## [71] 32.313700 36.765217 36.980224 36.564792 39.019192 37.469823 37.225302
## [78] 37.628163 41.304755 36.682976 41.735017 39.553862 42.977428 41.887912
## [85] 41.980186 39.545399 40.852016 44.987008 44.958597 48.834057 47.227912
## [92] 42.122757 43.955537 46.449343 44.850576 47.175137 48.671894 49.193177
## [99] 53.400262 51.074369
```

```
plot(xs)
```



```
acf(xs, lag.max = 25, type = c("correlation"))
```

Series xs



Voire un petit souci. Normalement la saisonnalité doit se manifester là. pk elle se manifeste pas ? Parce quelle est absorbé par la tendance (il faut diminuer le coef sur le 0.5t)